

Copyright (c) 1993 - 2005 Compugen Ltd.	Gencore version 5.1.6				
OM nucleic - nucleic search, using sw model					
Run on:	September 18, 2005, 09:12:34 ; search time 1914.49 Seconds				
Scoring table:	IDENTITY_NUC (without alignments)				
Total number of hits satisfying chosen parameters:	1033.875 Million cell updates/sec				
Minimum DB seq length:	0				
Maximum DB seq length:	2000000000				
Post-processing:	Minimum Match 0% Maximum Match 100%				
database :	EST:*				
	1: gb_est1:*				
	2: gb_est2:*				
	3: gb_htc:*				
	4: gb_est3:*				
	5: gb_est4:*				
	6: gb_est5:*				
	7: gb_est6:*				
	8: gb_gsb1:*				
	9: gb_gsb2:*				
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.	SUMMARIES				
Result No.	Score	Query Length	DB ID	Description	ALIGNMENTS
c 1	29.4	56.5	463 7 N59414	LOCUS N59414 YN51e08_s1 DEFINITION YN51e08_s1 Soares fetal liver spleen INFSL mRNA sequence. IMAGE:246278 3, mRNA sequence.	RESULT 1
c 2	29.2	56.2	601 9 CL376379	ACCESSION N59414	N59414/c
c 3	28.8	55.4	415 2 BB223173	VERSION GI:12033304	
c 4	28	53.8	323 BP745682	KEYWORDS EST.	
c 5	27.8	53.5	421 2 BB580731	SOURCE ORGANISM Homo sapiens (human)	
c 6	27.8	53.5	742 8 AQ47941	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
c 7	27.6	53.1	563 2 BB240791	REFERENCE 1. (bases 1 to 463)	
c 8	26.8	51.9	630 9 CB825712	AUTHORS Hillier,L., Lennon,G., Becker,M., Ronald,M.F., Chiapelli,B., Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,B., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.	
c 9	26.8	51.5	212 7 CP588249	TITLE Generation and analysis of 280,000 human expressed sequence tags	
c 10	26.9	51.5	220 9 BX945244	JOURNAL Genome Res. 6 (9), 807-828 (1996)	
c 11	26.8	51.5	471 8 BB421419	JOURNAL 97044478	
c 12	26.8	51.5	764 9 CNS05200	MEDLINE 8889549	
c 13	26.8	51.5	793 8 BB417682	COMMENT Contact: Wilson RK	
c 14	26.8	51.5	814 8 BB612324	Washington University School of Medicine	
c 15	26.8	51.5	864 9 CG110241	Washington University School of Medicine	
c 16	26.8	51.5	1035 9 CM50683	Al311553 Tetradon	
c 17	26.8	51.5	1037 9 CL992336	BZ4241553 Tetradon	
c 18	26.8	51.5	1962 9 AG3390999	BZ441562 INFSL	
c 19	26.6	51.2	544 8 BB21511	BZ21511 CH23.0-249	
c 20	26.6	51.2	566 4 BB528842	BZ528842 BZ528942	
c 21	26.6	51.2	586 9 CL307851	CL307851 02S0155-1	
c 22	26.6	51.2	600 7 CN747739	CN747739 SALUS008	
c 23	26.6	51.2	672 8 AZ062088	AZ062088 RRCI-23-4	
c 24	26.6	51.2	768 9 AG404371	AG404371 Mus muscu	

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## Om nucleic - nucleic search, using bw model

Run on: September 18, 2005, 02:55:23 ; Search time 239.707 Seconds

1284.177 Million cell updates/sec

Title: US-10-798-192-4

Perfect score: 52

Sequence: 1 ggtgatcacaaagacatg.....atttactgttttatggc 52

Scoring table: IDENTITY\_NUC

Gapop 10<sup>-0</sup>, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*

1: geneseqn1980s:\*

2: geneseqn1990s:\*

3: geneseqn200s:\*

4: geneseqn2001ab:\*

5: geneseqn2001bb:\*

6: geneseqn2002ab:\*

7: geneseqn2002bb:\*

8: geneseqn2003ab:\*

9: geneseqn2003bb:\*

10: geneseqn2003cb:\*

11: geneseqn2004ab:\*

12: geneseqn2004bb:\*

13: geneseqn2004cb:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
C 1	44	84.6	307 10 ADF2898	RESULT 1 ID ADF42898/C XX ADF42898 standard; DNA, 307 BP.
C 2	44	84.6	4404 10 ADL13987	XX AC ADF42898;
C 3	44	84.6	4404 10 ADL13987	XX DT 12-FEB-2004 (first entry)
C 4	44	84.6	4652 3 AAZ4209	XX DE AAV5 cap protein derived intron DNA.
C 5	44	84.6	4652 6 AAS17712	XX KW helper construct; packaging; Cap protein; serotype A; serotype B; intron; P40 promoter; P5 promoter; P19 promoter; Rep protein; gene therapy; ds.
C 6	59.5	2307 3 AAZ49212	XX KW P40 promoter; P5 promoter; P19 promoter; Rep protein; gene therapy; ds.	
C 7	52.7	5041 4 AAH18245	XX OS Adeno-associated virus 5.	
C 8	51.2	2477 4 AAH17555	XX OS WO2003074685,A1.	
C 9	50.8	203700 13 ABD3368	XX PD 12-SEP-2003.	
C 10	50.8	243390 13 ABD3366	XX PR 07-MAR-2003; 2003WO-EP002351.	
C 11	50.4	404 4 AAH191464	XX PA (MEDI-) MEDIGENE AG.	
C 12	49.6	404 4 ADR02848	XX PI Hoerer M, Dubielzig R, Ries S, Kosfeld-Bergauer M;	
C 13	49.6	2736 11 ADI31648	XX DR WPI; 2003-690032/65.	
C 14	49.6	3090 2 AAB085925	XX PT New helper construct for packaging adeno-associated virus, useful for preparing viral particles for gene therapy, encodes Cap and Rep genes of different serotypes.	
C 15	49.6	3492 6 AB167678	XX PS Disclosure; Fig 10; 43pp; German.	
C 16	49.6	3492 8 ABD52788	XX XX This invention describes a novel helper construct for packaging recombinant adeno-associated virus (rAAV) in which the sequence encoding the Cap protein of serotype A is linked to regulatory regions of a serotype B AAV, the Cap protein intron is of Serotype A and the sequences encoding Cap and Rep are functionally separate. The invention also describes a system for packaging rAAV comprising a helper construct that encodes a functional Rep from a serotype other than A, particularly B and a host cell for packaging rAAV of serotype A containing a copy of the	
C 17	49.6	80321 11 ACM45012	CC CC	
C 18	49.6	492 4 AB125255	CC CC	
C 19	49.2	3732 4 ABL25254	CC CC	
C 20	49.2	135827 13 ABD33219	CC CC	

AAC61741 Gene seqeu Adk64037 Disease t Abi12736 Drosophil Abi17820 P40V5 RN Acn4490 Human gen Aa157272 BA38B23- Aai95559 Human neu Aaa46343 Tumour su Abi31250 Signal tr Aak65236 Human imm Abz73979 Secreted Abi16919 Human sec Abz67560 Human sec Aak65235 Human imm Abz73980 Secreted Abi16920 Human sec Abz67561 Human sec Aab8766 Human imm Adq97173 Human can Adq97316 Mouse can Aab15519 Human tyr Abd47900 Human tra Aab161371 Soybean 2 Aab161372 Soybean 2 Abi14118 Drosophil

Copyright (c) 1993 - 2005 Compugen Ltd.	GenCore version 5.1.6	
Run on:	September 18, 2005, 09:12:34 ; Search time 1914.49 Seconds	
Scoring table:	IDENTITY_NUC	
Gapop:	10.0 , Gapext 1.0	
Searched:	34239544 seqs, 19032134700 residues	
Total number of hits satisfying chosen parameters:	68479086	
Minimum DB seq length:	0	
Maximum DB seq length:	2000000000	
Post-processing: Minimum Match 0%		
Maximum Match 100%		
listing first 45 summaries		
Database :		
EST:*		
1: gb_est1:*		
2: gb_est2:*		
3: gb_ntc:*		
4: gb_est3:*		
5: gb_est4:*		
6: gb_est5:*		
7: gb_est6:*		
8: gb_gbs1:*		
9: gb_gbs2:*		
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.		
SUMMARIES		
Result No.	Query Match Length DB ID	Description
C 1	29.2 56.2 601 9	CL376379 RPCI44_44
C 2	28.8 55.4 415 2	BB223173 kp7f55_y
C 3	28.4 54.6 463 7	NS9414 Yv5te08_s1
C 4	27.8 53.3 421 2	BB580731 kp1a01_y
C 5	27.8 53.5 742 8	AQ479841 RPCI-11-2
C 6	27.6 53.1 323 5	BPF45682 BP745682
C 7	27.6 53.1 563 2	BB240791 RST40840
C 8	27.5 51.9 630 9	CB825712 tigr-3ba-
C 9	26.8 51.5 220 9	BK945244 Arabidops
C 10	26.8 51.5 471 8	B242119 lf80hs5_p
C 11	26.8 51.5 764 9	AL317553 Tetradon
C 12	26.8 51.5 793 8	B2417682 iF80hs5_p
C 13	26.8 51.5 814 9	B2612324 WHABV4TR
C 14	26.8 51.5 864 9	CG110241 PRTB040TB
C 15	26.8 51.5 1035 9	CW509683 ZMMBC000
C 16	26.8 51.5 1037 9	CL992336 ZMMBE000
C 17	26.8 51.5 1962 9	AG390999 Mus musculus
C 18	26.6 51.5 566 4	B0528842 B0528842
C 19	26.6 51.2 600 7	CNT47739 SAL US008
C 20	26.6 51.2 1005 7	CK282807 EST745529
C 21	26.4 50.8 212 7	CF588249 USDA_FP1
C 22	26.4 50.8 794 8	BRH01231 Sm1-5412
C 23	26.4 50.8 856 7	CN168256 AGENCOURT
C 24	26.4 50.8 888 5	BU410024 603159192
FEATURES	Location/Qualifiers	
source	1. -601 /organism="Sub scrofa" /mol_type="genomic DNA" /strain="four pigs (breed: 37.5% Yorkshire Landrace and 25% Meishan)" /db_xref="taxon:9833" /clone="RPCI44_444113"	
ALIGNMENTS		

GenCore version 5.1.6  
copyright (c) 1993 - 2005 Compugen ltd.

OM nucleic - nucleic search, using sw model  
Run on: September 18, 2005, 06:10:24 ; Search time 506.39 Seconds  
(without alignments) 2679.250 Million cell updates/sec

Title: US-10-798-192-1  
Perfect score: 28  
Sequence: 1 ggcacataatgatttaaacaggatg 28

Scoring table: IDENTITY\_NUC  
Gapp 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenBank:  
1: gb\_ba: \*  
2: gb\_hig: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sis: \*  
12: gb\_sp: \*  
13: gb\_un: \*  
14: gb\_vl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description	ORIGIN	RESULT 1	RESULT 2
1	26.4	94.3	2370	6 AX753285 Sequence	AX753285	AX753285 Sequence	AX753277 Sequence
2	26.4	94.3	2489	6 AX753280 Sequence	AX753280	AX753280 Sequence	AX753288 Sequence
3	26.4	94.3	2495	6 AX753281 Sequence	AX753281	AX753281 Sequence	AX75289 Sequence
4	26.4	94.3	2501	6 AX753275 Sequence	AX753275	AX753275 Sequence	AX753287 Sequence
5	26.4	94.3	2504	6 AX753278 Sequence	AX753278	AX753278 Sequence	AX753290 Sequence
6	26.4	94.3	2969	6 AX753303 Sequence	AX753303	AX753303 Sequence	AX75286 Sequence
7	26.4	94.3	3084	6 AX753284 Sequence	AX753284	AX753284 Sequence	AX753272 Sequence
8	26.4	94.3	3098	6 AX753268 Sequence	AX753268	AX753268 Sequence	AX753273 Sequence
9	26.4	94.3	3098	6 AX753254 Sequence	AX753254	AX753254 Sequence	AX753291 Sequence
10	26.4	94.3	3098	6 AX753255 Sequence	AX753255	AX753255 Sequence	AX753292 Sequence
11	26.4	94.3	3098	6 AX753282 Sequence	AX753282	AX753282 Sequence	AX753294 Sequence
12	26.4	94.3	3102	6 AX753262 Sequence	AX753262	AX753262 Sequence	AX753259 Sequence
13	26.4	94.3	3105	6 AX753264 Sequence	AX753264	AX753264 Sequence	AX753274 Sequence
14	26.4	94.3	3105	6 AX753265 Sequence	AX753265	AX753265 Sequence	AX753283 Sequence
15	26.4	94.3	3105	6 AX753266 Sequence	AX753266	AX753266 Sequence	AX753250 Sequence
16	26.4	94.3	3106	6 AX753261 Sequence	AX753261	AX753261 Sequence	AX753249 Sequence
17	26.4	94.3	3106	6 AX753263 Sequence	AX753263	AX753263 Sequence	AX753249 Sequence
18	26.4	94.3	3106	6 AX753279 Sequence	AX753279	AX753279 Sequence	AX753251 Sequence
19	26.4	94.3	3113	6 AX753276 Sequence	AX753276	AX753276 Sequence	AP513851 Adeno-abb

## ALIGNMENTS

Qry	1	GAGCAATAATGATTAAACCGAGTAG	28
Db	819	GAGCAATAATGATTAAACCGAGTAG	846
Source	unidentified		
Organism	unidentified		
Reference	1		
Authors	Gao,G., Wilson,J.M. and Alvira,M.		
Title	A method of detecting and/or identifying adeno-associated virus (AAV) sequences and isolating novel sequences identified thereby		
Journal	Patent: EP 1310571-A 40 14-MAY-2003; The Trustees of The University of Pennsylvania (US)		
Features	/location/qualifiers		
Source	1. 2370 /organism="unidentified" /mol-type="unassigned DNA" /db_xref="taxon:3264" /notes="new AAV serotype, clone 43.5"		

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: September 18, 2005, 06:10:24 ; Search time 578.732 Seconds  
(without alignments)  
2679.250 Million cell updates/sec

Title: US-10-798-192-2  
Perfect score: 32  
Sequence: 1 getcttagaccgatgcgtttatcg 32

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

#### Database :

GenBank:\*

- 1: gb\_ba:\*
- 2: gb\_hgt:\*
- 3: gb\_in:\*
- 4: gb\_em:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

#### ALIGNMENTS

#### RESULT 1

BD242766/c

DEFINITION Adeno-associated virus serum type 1 nucleic acid sequence, vector

LOCUS BD242766

ACCESSION BD242766

VERSION GI:33052536

KEYWORDS JP:2002529098-A/1.

SOURCE unidentified

ORGANISM unclassified

REFERENCE 1 (bases 1 to 4718)

AUTHORS Wilson,J.M. and Xiao,W.

TITLE

JOURNAL

COMMENT THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA

OS AAV-1

PN JP:2002529098-A/1

PD 10-SEP-2002

PP 02-NOV-1999 JP 2000591227

PR 05-NOV-1998 US 60101714

PI JAMES M WILSON WEIDONG XIAO

PC C12N15/09,A61K31/71,A61K48/00,A61P43/00,C12N1/15,C12N1/19, PC

PC C12N1/21, PC

PC C12N5/10,C12P21/02/A61K35/76,C12N5/00,C12N5/00 CC

Adeno-associated virus serum type 1 nucleic acid sequence, CC

AL355931 Neuropor

AC110671 Mus muscu

AC150411 Branchios

AL593062 Zebrafish

Continguation (3 of

AC150643 Bos tauru

AC129120 Rattus no

AC135669 Rattus no

AC037745 Rattus no

AL6633080 Mouse DNA

AC087256 Homo sapi

AC021868 Homo sapi

AC13775 Mus muscu

AC11524 Rattus no

AC019209 Homo sapi

Query Match 78.1%; Score 25; DB 6; Length 4718;  
Best Local Similarity 100.0%; Pred. No. 0.51; 0; Indels 0; Gaps 0;

Matches 25; Conservative 0; Mismatches 0;